DOCUMENT RESUME

ED 061 287 TM 001 238

AUTHOR Williams, John D.; And Others
TITLE A Comparison of Raw Gain Scores, Residual Gain Scores, and the Analysis of Covariance with Two Modes of Teaching Reading.
PUB DATE Apr 72
NOTE 17p.; Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, Illinois, April 1972

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS *Analysis of Covariance; *Comparative Statistics; Correlation; Data Analysis; Elementary School Students; *Grouping (Instructional Purposes); Inservice Teaching; Mathematical Models; Nongraded Classes; Post Testing; Predictor Variables; Pretesting; *Raw Scores; *Reading Instruction; Reading Tests; Rural Schools; Scores; Teaching Techniques
IDENTIFIERS *Attitudes Toward Reading Inventory; California Reading Test

ABSTRACT

Two methods of reading instruction (homogeneous grouping and graded classes) are compared for 165 students in 8 rural North Dakota schools by raw gain scores, residual gain scores, and the analysis of covariance. (CK)

ED 061287

A Comparison of Raw Gain Scores, Residual Gain Scores, and the Analysis of Covariance with Two Modes of Teaching Reading

John D. Williams
The University of North Dakota

Roger T. Maresh
Elementary Principal-Coordinator
Rockwood School District
Eureka, Missouri

James D. Peebles
Director, Diagnostic Reading Program
University City, Missouri Public School System

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

TM 001 238

The measurement of change has been seen to be one of the most difficult issues in psycho-educational research (Harris, 1963). Several different solutions have been proposed, and almost simultaneously, have been criticized. When pre and post-testing have taken place, an intuitively pleasing approach has been the use of raw gains (that is, the post-test score minus the pre-test score for each subject). The use of this measure has been severely criticized. Ruch (1970) has indicated his displeasure with gain scores because of this disregard for the psychology of learning. Because learning, in its latter phases is often characterized by a negatively accelerated curve, those students who enter an experiment with more practice in the skill or concept being tested will be handicapped by the gain score approach. The student who has a smaller amount of prior practice enters the experiment during the initial phase of learning, which will allow him to be in a period of rapid acceleration in regard to measured learning.

A common approach to the problem of measuring change when a pre and post-test have been used is the analysis of covariance. The analysis of covariance is often used when the assignment of subjects to an experiment has been made on some basis other than strict randomization. The analysis of covariance takes into account the correlation between the pre-test and the post-test. More specifically, it is helpful to look at the process of the analysis of covariance as it can be generated through the use of linear models. Because the present application is concerned with two modes of instruction, one mode being the vertically grouped method of teaching reading, and the second method being the more typical graded method of teaching reading, and because a pre- and post-test are being used, the linear models developed here will represent that specific situation.

First, a <u>full model</u> can be defined. A full model is essentially a model that contains all the information relevant to the data analysis. For this

specific situation, the full model is:

$$Y = b_0 + b_1X_1 + b_2X_2 + e_1, \qquad (1)$$

where

Y = the post-test score,

X_1 = the pre-test score,

X_2 = 1 if the score is from a member of the vertical group; 0 otherwise,

b_0 = the Y-intercept,

b_1 = the regression coefficient for X_1,

b_2 = the regression coefficient for X_2, and

e_1 = the error in prediction with the full model.

If this model is solved using a multiple linear regression computer programming routine, part of the output includes the multiple correlation coefficient (R). For the present usage, since a full model is being used, the R value found from the use of equation 1 will be labeled R_{FM}.

Similarly, a <u>restricted model</u> can be developed, using the pre-test as the predictor variable:

$$Y = b_0 + b_1X_1 + e_2, \qquad (2)$$

where

Y = the post-test score,

X_1 = the pre-test score,

b_0 = the Y-intercept (the b_0 value for equation 2 will, in general, be different from the b_0 value in equation 1),

b_1 = the regression coefficient for X (again, the b_1 value for equation 2 will, in general, be different from the b_1 value found in equation 1), and

e_2 = the error in prediction with the restricted model.

The restricted model will also yield an R value, and it will be labeled R_{RM}. The F test for the analysis of covariance is given by:

$$F = \frac{(R^2_{FM} - R^2_{RM}) / 1}{(1 - R^2_{FM}) / N - 3} . \qquad (3)$$

This F test is specific for this situation. A more general F test would be given by:

$$F = \frac{(R^2_{FM} - R^2_{RM})/(k - 1)}{(1 - R^2_{FM})/(N - C - k)} \qquad (4)$$

where

k is the number of groups,

N is the number of subjects, and

C is the number of covariates.

It is also possible to find adjusted means for the analysis of covariance.

DuBois (1957, 1970) has worked extensively with the residual gain analysis. Essentially, the residual gain analysis can be conceptualized as a part correlation between the group membership variable(s) and the residual in the post-test data when using the pre-test as the predictor. As a model, this can be accomplished easily in two stages with an ordinary multiple regression program. The first model is:

$$Y = b_0 + b_1 X_1 + e_3, \qquad (5)$$

where

Y = the post-test score,

X_1 = the pre-test score,

b_0 = the Y-intercept (the value for b_0 in equation 5 will, in general, be different than previously defined b_0 values),

b_1 = the regression coefficient for X_1 (the value for b_1 in equation 5 will, in general, be different than previously defined b_1 values), and

e_3 = the error in prediction for this model.

The focus in the residual gains analysis is on the residual errors (e_3) for each subject. These residual errors become the criterion scores, and the group membership variable(s) are used to complete the residual gain analysis. The model is as follows:

$$Y' = b_0 + b_2 X_2 + e_4, \qquad (6)$$

where

Y' = the residual errors found from the use of equation 5,

X_2 = 1 if the score is from a member of the vertical group; 0 otherwise,

b_0 = the Y-intercept (the b_0 value in equation 6 will, in general, be different than the previous b_0 values),

b_2 = the regression coefficient for X_2 (the b_2 value in equation 6 will, in general, be different from the b_2 value in equation 2), and

e_4 = the error in prediction for this model.

The use of the residual gain analysis has been based upon the following considerations: the residual gain scores will be uncorrelated with initial status, whereas it can be expected that the raw gain scores will show a negative correlation with initial states; whenever all subjects do not start at a common point (so that the methods of common points of mastery could not be used), the residual score nevertheless:

1. can be defined precisely and accurately,
2. the residual does not require the use of a ratio scale to measure initial and final states, and
3. higher ordered residual gains can be found.

Carver (1970) has compared the residual gain analysis to the method of common points of mastery, initially proposed by Ruch (1936). Conceptually, both of

these measures were employed to overcome the difficulties involved with the raw gain scores. Employing both methodologies on empirical data, Carver was able to find only moderate correlations between the measures.

Subjects

The subjects for this study included 165 students in 8 rural North Dakota schools. All the students were enrolled in learning situations where the instructor was an intern (or in some cases, graduates of the New School program) from the New School of the University of North Dakota, an experimental program funded by the United States Office of Education. The vertically grouped subjects were those students who were enrolled in a classroom setting that allowed a non-graded approach to instruction in several areas. Thus, the reading instruction took place in a homogeneous group rather than an age (or graded) group. The second group of students received their reading instruction in a graded group (i.e., Grade Four, Grade Five, etc). The grade levels involved were Grades Two through Grade Six.

Method

Two instruments were administered on a pre and post-test basis. Pre-tests were administered in October, 1970, and post-tests were administered in May, 1971. The vocabulary and comprehension sections of the California Reading Test (Tiegs and Clark, 1957) was used at all five grade levels. The Attitudes Toward Reading Inventory (Hunt, 1961) was used with only grades four, five, and six. The Attitudes Toward Reading Inventory has two subtests, Attitudes Toward Reading, and Attitudes Toward Reading Class.

Results

Tables 1-6 show the analysis of the data. Each table includes means for the pre-test and post-test, adjusted means, raw gain, and residual gain for the two modes of instruction in reading. Included also are the F values, R, R^2, and

SS_T (Sum of squares total) for each analysis. This method of presentation is used for economy of space and to allow for ease in comparing the different results. Actually, a summary table could be generated for all five different sets of data analyses. In the following tables the R value is the correlation between the dichotomous predictor (group membership) and the criterion scores, with the exception of the analysis of covariance (illustrated here under the name adjusted means), which is completed as it was described earlier. While there are different approaches to measuring the strength of relationships with dichotomous information, using Walberg's (1971) approach, the R^2 value is interpreted as being the amount of criterion variance accounted for by group membership. Also included in each table is some indication of significance. There is a slight discrepancy with the analysis of covariance (adjusted means) and the residual gains analysis. The degrees of freedom for the analysis of covariance and the residual gains analysis in this situation will actually be one less than the degrees of freedom listed under each table. In that no interpretations are changed in the present situation in regard to the differences in degrees of freedom, that slight difference in degrees of freedom is not indicated in the tables.

TABLE 1

SUMMARY DATA RELATING TO SECOND GRADE VOCABULARY SCORES

Vocabulary Scores - Grade 2 (N = 35)

	Pre-test	Post-test	Adjusted Mean	Raw Gain	Residual Gain
Vertical Group	2.359	3.194	3.229	.835	-.022
Graded Group	2.611	3.306	3.273	.694	.021
$F = t^2$	1.840	.695	.119	.581	.112
R	.230	.144	Full .392 Rest .388	.132	.059
R^2	.053	.021	Full .154 Rest .151	.014	.003
SS_T	10.535	5.267	4.475	10.022	4.474

Critical value for significance at .05 level with df = 1, 33 is 4.14.
Critical value for significance at .01 level with df = 1, 33 is 7.47.

TABLE 2

SUMMARY DATA RELATING TO THIRD GRADE VOCABULARY SCORES

Vocabulary Scores - Grade 3 (N = 48)

	Pre-test	Post-test	Adjusted Mean	Raw Gain	Residual Gain
Vertical Group	3.667	4.270	4.300	.603	-.049
Graded Group	3.789	4.483	4.433	.694	.082
$F = t^2$	.701	2.246	1.537	.603	1.512
R	.123	.216	Full .688 Rest .675	.114	.180
R^2	.015	.047	Full .473 Rest .456	.013	.032
SS_T	11.192	11.000	5.989	7.212	5.988

Critical value for significance at .05 level with df = 1, 46 is 4.05.
Critical value for significance at .01 level with df = 1, 46 is 7.21.

TABLE 3

SUMMARY DATA RELATING TO FOURTH GRADE VOCABULARY SCORES

Vocabulary Scores - Grade 4 (N = 37)

	Pre-test	Post-test	Adjusted Mean	Raw Gain	Residual Gain
Vertical Group	5.244	5.900	5.748	.656	-.134
Graded Group	4.986	5.876	5.992	.890	.102
$F = t^2$	1.161	.005	1.328	1.290	1.283
R	.179	.012	Full .771 Rest .761	.189	.191
R^2	.032	.001	Full .594 Rest .579	.036	.036
SS_T	18.829	33.243	14.015	14.016	14.014

Critical value for significance at .05 level with df = 1, 35 is 4.12.
Critical value for significance at .01 level with df = 1, 35 is 7.42.

TABLE 4

SUMMARY DATA RELATING TO FIFTH GRADE VOCABULARY SCORES

Vocabulary Scores - Grade 5 (N = 27)

	Pre-test	Post-test	Adjusted Mean	Raw Gain	Residual Gain
Vertical Group	5.880	6.580	6.536	.700	.120
Graded Group	5.800	6.318	6.344	.518	-.071
$F = t^2$	.064	.599	.954	.869	.952
R	.050	.153	Full .829 Rest .821	.183	.195
R^2	.003	.023	Full .687 Rest .674	.033	.038
SS_T	15.816	18.534	6.038	6.234	6.038

Critical value for significance at .05 level with df = 1, 25 is 4.24.
Critical value for significance at .01 level with df = 1, 25 is 7.77.

TABLE 5

SUMMARY DATA RELATING TO SIXTH GRADE VOCABULARY SCORES

Vocabulary Scores - Grade 6 (N = 28)

	Pre-test	Post-test	Adjusted Mean	Raw Gain	Residual Gain
Vertical Group	6.356	7.089	7.162	.774	.033
Graded Group	6.479	7.147	7.113	.669	-.016
$F = t^2$	.153	.027	.044	.0799	.045
R	.077	.032	Full .771 Rest .770	.055	.042
R^2	.006	.001	Full .594 Rest .593	.003	.002
SS_T	15.847	20.297	8.248	8.507	8.248

Critical value for significance at .05 level with df = 1, 26 is 4.22.
Critical value for significance at .01 level with df = 1, 26 is 7.72.

TABLE 6

SUMMARY DATA RELATING TO SECOND GRADE COMPREHENSION SCORES

Comprehension Scores - Grade 2 (N = 35)

	Pre-test	Post-test	Adjusted Mean	Raw Gain	Residual Gain
Vertical Group	2.047	3.094	3.216	1.047	.057
Graded Group	2.550	3.194	3.079	.644	-.054
$F = t^2$	7.93**	.318	.594	4.838*	.478
R	.440	.098	Full .481 Rest .466	.358	.121
R^2	.194	.010	Full .231 Rest .217	.128	.015
SS_T	11.419	9.228	7.226	11.084	7.225

*Significant at .05 level. Critical value for significance at .05 level with df = 1, 33 is 4.14.
**Significant at .01 level. Critical value for significance at .01 level with df = 1, 33 is 7.47.

TABLE 7

SUMMARY DATA RELATING TO THIRD GRADE COMPREHENSION SCORES

Comprehension Scores - Grade 3 (N = 48)

	Pre-test	Post-test	Adjusted Mean	Raw Gain	Residual Gain
Vertical Group	3.678	4.307	4.321	.627	-.035
Graded Group	3.744	4.439	4.415	.694	.058
$F = t^2$	.203	.890	.696	.308	.693
R	.066	.138	Full .624 Rest .617	.082	.123
R^2	.004	.019	Full .389 Rest .381	.007	.015
SS_T	9.548	10.358	6.419	7.780	6.419

Critical value for significance at .05 level with df = 1, 46 is 4.05.
Critical value for significance at .01 level with df = 1, 46 is 7.21.

TABLE 8

SUMMARY DATA RELATING TO FOURTH GRADE COMPREHENSION SCORES

Comprehension Scores - Grade 4 (N = 37)

	Pre-test	Post-test	Adjusted Mean	Raw Gain	Residual Gain
Vertical Group	5.237	6.244	6.205	1.006	.189
Graded Group	5.176	5.843	5.872	.667	-.144
$F = t^2$	.046	1.160	2.719	2.847	2.714
R	.036	.179	Full .850 Rest .837	.274	.272
R^2	.001	.032	Full .723 Rest .701	.075	.074
SS_T	25.910	45.490	13.616	13.923	13.614

Critical value for significance at .05 level with df = 1, 35 is 4.12.
Critical value for significance at .01 level with df = 1, 35 is 7.42.

TABLE 9

SUMMARY DATA RELATING TO FIFTH GRADE COMPREHENSION SCORES

Comprehension Scores - Grade 5 (N = 27)

	Pre-test	Post-test	Adjusted Mean	Raw Gain	Residual Gain
Vertical Group	6.330	7.070	6.626	.740	.149
Graded Group	5.394	6.053	6.314	.659	-.087
$F = t^2$	8.124**	10.778**	2.031	.159	1.501
R	.495	.549	Full .864 Rest .851	.079	.243
R^2	.245	.301	Full .747 Rest .724	.006	.059
SS_T	22.485	21.617	5.954	6.567	5.952

Critical value for significance at .05 level with df = 1, 25 is 4.22.
**Significant at .01 level. Critical value for significance at .01 level with df = 1, 25 is 7.77.

TABLE 10

SUMMARY DATA RELATING TO SIXTH GRADE COMPREHENSION SCORES

Comprehension Scores - Grade 6 (N = 28)

	Pre-test	Post-test	Adjusted Mean	Raw Gain	Residual Gain
Vertical Group	6.367	7.378	7.512	1.011	.201
Graded Group	6.616	7.274	7.210	.658	-.095
$F = t^2$	.598	.101	2.094	2.746	2.043
R	.150	.062	Full .786 Rest .765	.309	.275
R^2	.023	.004	Full .618 Rest .585	.095	.076
SS_T	16.864	17.138	7.096	7.978	7.097

Critical value for significance at .05 level with df = 1, 26 is 4.22.
Critical value for significance at .01 level with df = 1, 26 is 7.72.

TABLE 11

SUMMARY DATA RELATING TO FOURTH GRADE
ATTITUDES TOWARD READING SCORES

Attitudes Toward Reading Scores - Grade 4 (N = 37)

	Pre-test	Post-test	Adjusted Mean	Raw Gain	Residual Gain
Vertical Group	24.50	26.000	25.816	1.500	.490
Graded Group	24.048	24.810	24.950	.762	-.374
$F = t^2$	.114	.755	.756	.492	.754
R	.057	.145	Full .706 Rest .698	.118	.147
R^2	.003	.021	Full .498 Rest .487	.014	.022
SS_T	570.808	610.105	312.619	356.755	312.615

Critical value for significance at .05 level with df = 1, 35 is 4.12.
Critical value for significance at .01 level with df = 1, 35 is 7.42.

TABLE 12

SUMMARY DATA RELATING TO FIFTH GRADE
ATTITUDES TOWARD READING SCORES

Attitudes Toward Reading Scores - Grade 5 (N = 27)

	Pre-test	Post-test	Adjusted Mean	Raw Gain	Residual Gain
Vertical Group	26.100	26.700	24.339	.600	.588
Graded Group	21.412	21.765	23.154	.354	.346
$F = t^2$	6.718*	6.232*	.586	.031	.460
R	.460	.447	Full .793 Rest .787	.035	.137
R^2	.212	.200	Full .629 Rest .619	.001	.019
SS_T	653.407	768.516	292.629	306.665	292.626

*Significant at .05 level. Critical value for significance at .05 level with df = 1, 25 is 4.24.
Critical value for significance at .01 level with df = 1, 25 is 7.77.

TABLE 13

SUMMARY DATA RELATING TO SIXTH GRADE
ATTITUDES TOWARD READING SCORES

Attitudes Toward Reading Scores - Grade 6 (N = 28)

	Pre-test	Post-test	Adjusted Mean	Raw Gain	Residual Gain
Vertical Group	24.444	25.000	24.175	.555	.197
Graded Group	22.053	23.474	23.864	.526	-.093
$F = t^2$	1.896	1.092	.064	.455	.060
R	.261	.201	Full .625 Rest .624	.131	.049
R^2	.068	.041	Full .391 Rest .389	.017	.002
SS_T	514.105	342.962	215.518	334.712	215.518

Critical value for significance at .05 level with df = 1, 26 is 4.22.
Critical value for significance at .01 level with df = 1, 26 is 7.72.

TABLE 14

SUMMARY DATA RELATING TO FOURTH GRADE
ATTITUDES TOWARD READING CLASS SCORES

Attitudes Toward Reading Class Scores - Grade 4 (N = 37)

	Pre-test	Post-test	Adjusted Mean	Raw Gain	Residual Gain
Vertical Group	38.375	40.188	39.450	1.813	.103
Graded Group	36.619	37.000	37.562	.381	-.78
$F = t^2$	1.312	2.862	1.474	.847	1.418
R	.190	.275	Full .641 Rest .621	.154	.200
R^2	.036	.076	Full .411 Rest .386	.024	.040
SS_T	774.702	20.670	750.256	787.997	750.254

Critical value for significance at .05 level with df = 1, 35 is 4.12.
Critical value for significance at .01 level with df = 1, 35 is 7.42.

TABLE 15

SUMMARY DATA RELATING TO FIFTH GRADE
ATTITUDES TOWARD READING CLASS SCORES

Attitudes Toward Reading Class Scores - Grade 5 (N = 27)

	Pre-test	Post-test	Adjusted Mean	Raw Gain	Residual Gain
Vertical Group	37.600	39.100	37.434	1.500	.819
Graded Group	33.882	35.000	35.980	1.118	-.482
$F = t^2$	2.947	4.289*	1.176	.076	1.046
R	.325	.383	Full .815 Rest .805	.055	.204
R^2	.106	.147	Full .664 Rest .648	.003	.042
SS_T	825.183	722.754	254.847	305.184	254.844

Critical value for significance at .05 level with df = 1, 25 is 4.24.
Critical value for significance at .01 level with df = 1, 25 is 7.77.

TABLE 16

SUMMARY DATA RELATING TO SIXTH GRADE
ATTITUDES TOWARD READING CLASS SCORES

Attitudes Toward Reading Class Scores - Grade 6 (N = 28)

	Pre-test	Post-test	Adjusted Mean	Raw Gain	Residual Gain
Vertical Group	35.667	36.000	36.726	.333	-.164
Graded Group	37.053	37.316	36.972	.263	.077
$F = t^2$	.556	.404	.025	.002	.025
R	.145	.124	Full .699 Rest .699	.009	.032
R^2	.021	.015	Full .489 Rest .489	.0001	.001
SS_T	560.677	690.678	353.465	381.713	353.462

Critical value for significance at .05 level with df = 1, 26 is 4.22.
Critical value for significance at .01 level with df = 1, 26 is 7.72.

Discussion

It should be abundantly clear from the 16 tables that the three approaches to psycho-educational change are different. While this set of data does not exhibit strong relationships between the dichotomous predictor and the various criteria, the use of the statistical significance approach would occasionally yield different interpretations. Perhaps the most objective comparison between the three measures would be the R^2 term (for the analysis of covariance, or adjusted means approach $R^2_{FM} - R^2_{RM}$). Only one significant difference is found in the three measures. In Table 6, the raw gain is significant ($p < .05$), but, under exactly the conditions that would tend to make this occur, the vertical group was significantly smaller than the graded group on the pre-test, but this difference was almost erased on the post-test. In terms of the raw gains score, this produced a significant difference in favor of the vertical group.

In general, the interpretations of the tests would be in the same direction, although the reverse is true in Table 1. In Table 1, the raw gain scores favor the vertical group, while the analysis of covariance (adjusted means) and the residual gain scores favor the graded group.

References

Carver, R. P. "The Relation between Two Measures of Learning: Residual Gain and Common Points of Mastery," in DuBois, P. H. and Mayo, G. D., Research Strategies for Evaluating Training. Chicago: Rand McNally, 1970, 100-106.

DuBois, P. H. Multivariate Correlational Analysis. New York: Harper, 1957.

DuBois, P. H. "Correlational Analysis in Training Research," in DuBois, P. H. and Mayo, G. D., Research Strategies for Evaluating Training. Chicago: Rand McNally, 1970, 109-116.

Harris, C. W. Problems in Measuring Change. Madison, Wisconsin: University of Wisconsin Press, 1963.

Hunt, L. Teacher and Pupil Attitude and Performance in Relation to Number of Books Used in First Grade Reading, Final Report, Project Number 3012, Contract No. OE-6-10, 029, U.S. Department of Health, Education, and Welfare, Office of Education. 1967.

Ruch, F. L. "The Method of Common Points of Mastery as a Technique in Human Learning Experimentation," Psychological Review, 1936, 43, 229-234.

Ruch, F. L. "Measuring Gain from a Common Point of Mastery," in DuBois, P. H. and Mayo, G. D., Research Strategies for Evaluating Training. Chicago: Rand McNally, 1970, 94-99.

Tiegs, E. W. and Clark, W. W. Manual, California Achievement Tests, Complete Battery, Monterey, California: California Test Bureau, 1957.

Walberg, H. J. "Generalized Regression Models in Educational Research," American Educational Research Journal, 1971, 71-92.